Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Cross-language Data on Five Types of Prosodic Focus

Martin Ho Kwan Ip, Anne Cutler

The MARCS Institute, Western Sydney University, Penrith South, NSW 2751, Australia
m.ip@westernsydney.edu.au, a.cutler@westernsydney.edu.au

Abstract

To examine the relative roles of language-specific and language-universal mechanisms in the production of prosodic focus, we compared production of five different types of focus by native speakers of English and Mandarin. Two comparable dialogues were constructed for each language, with the same words appearing in focused and unfocused position; 48 speakers recorded two dialogues each in their respective native language. Duration, F_0 (mean, maximum, range), and rms-intensity (mean, maximum) of all critical word tokens were measured. Across the different types of focus, cross-language differences were observed in the degree to which English versus Mandarin speakers use the different prosodic parameters to mark focus, suggesting that while prosody may be universally available for expressing focus, the means of its employment may be considerably language-specific.

Index Terms: focus, language-specificity, English, Mandarin

1. Introduction

Information structure is a linguistic universal. As long as speech is used for communication between people, utterances will concern some things that are, in one sense or another [1], more important and some that are less important. As a result, all speakers have the option to convey this structure in the way they speak, and they may use prosody to do it. Indeed, Bolinger [2] listed the highlighting of more important elements as one of only two identifiable prosodic universals.

How this highlighting – expression of focus – is achieved by means of prosody has been shown, however, to differ across languages. For instance, cross-language experiments comparing speakers of English, French, and German have revealed language-specific strategies where only German speakers used duration cues to enhance new information [3]. At the same time, how prosodic focus is realised can also depend on the particular morpho-syntactic structure of the language, as in Wolof [4], where morphological markers for focus are available, so that speakers then do not redundantly highlight semantic salience by the use of prosodic cues. Given the variation across languages in the resources for marking focus in speech production, there may in consequence be no universal manner in which prosodic focus is processed in speech perception.

On the other hand, language-specific aspects of speech processing in perception may be based on common underlying mechanisms. For instance, prosody may be universally available as a resource that all speakers can use, to a varying degree, to highlight focus. Consistent with this view, experimental evidence has shown that speakers can indeed employ prosodic cues when their language has other means to signal focus. Consider the case of Mandarin Chinese and Japanese: In both of these languages, focus can also be marked via focus-sensitive particles or phrasing, but speakers have been shown to effect focus using prosodic parameters such as pitch and duration [5, 6, 7]. Moreover, both languages manage to employ prosody for focus expression in ways that do not interfere with the production of lexical tones or pitch accents (e.g., by expansion of pitch register). Therefore, prosody may play a universal role as a medium through which speakers can express focus, even though its use may vary widely – from languages where prosody is largely ignored for this purpose, to languages where it is the only way focus is expressed, with many cross-language differences in the precise way in which the parameters of prosody are used for this purpose.

Although the literature thus currently provides data about speakers' production of cues to focus in many languages, the experimental designs and the structure of the materials used in the existing studies are often quite different. It is therefore not always easy to reach conclusions from the existing results about universality and language-specificity in prosodic processing. The experiment we report here forms part of a larger cross-language project examining universal versus language-specific components of prosodic processing in English and Mandarin. In the present production component, we aimed to produce a substantial database of focused and unfocused realisations of the same words by multiple speakers in contexts that were both relatively realistic and closely matched across the two languages.

Our study further compares different expressions of focus. This allows us to address issues in semantic theory concerning whether focus is a unitary construct. According to Rooth [8], and more recently Krifka and Musan [9], there are no principled differences between different focus types, on the grounds that all expressions of focus evoke two semantic representations: the actual meaning of a focused expression and a set of alternatives. However, experimental evidence has revealed different acoustic correlates of focus in different contexts. For instance, Ouyang and Kaiser [10] found that Mandarin words produced with corrective focus had longer durations, greater pitch expansion, and larger intensity ranges, while words that indicated new information only showed duration lengthening and pitch expansion. Similarly, in English, speakers are more likely to produce rising (L+H*) accents to encode elements in the discourse that signal explicit contrastive information [11]. Although most previous studies only looked at two groups of focus, we here compare cases of focus across five different pragmatic contexts.

In this experiment, we test whether English and Mandarin speakers differ in the degree to which they use the various prosodic parameters to signal each type of focus. Speakers of languages with different prosodic systems may differ in their reliance on the different prosodic parameters, even when both languages realise prosodic focus in the same way.

 doi: 10.21437/SpeechProsody.2016-68

Table 1. *Examples of focused tokens by focus type in English and Mandarin. Note in the actual experiment, participants were given all dialogues in plain form.*

English	Mandarin
Wh-focus: Vendor: What are you after? Buyer: I'm after a **[SWEATER]**.	*Wh-focus*: 小贩: 你想买什么呀? 顾客: 我想买件**[毛衣]**。
Corrective focus: Inspector: You heard two books dropped? Student: No, I heard two **[GUNSHOTS]**.	*Corrective focus*: 警察: 你突然间听到两声炮响? 学生: 不是，我听到**[枪响]** 。
Confirmatory focus: Inspector: And there was more than one gunshot? Student: Yes that's right, I heard **[TWO]** gunshots.	*Confirmatory focus*: 警察: 而且还不只一次枪响，对吗? 学生: 是啊没错，我听到**[两]** 声枪响。
Parallel focus: Buyer: I want to buy a **[GREEN]** sweater for my friend and a **[RED]** sweater for ……	*Parallel focus*: 顾客: 我要买件**[绿色]**的毛衣给我朋友, **[蓝色]**的毛衣给…
New-information focus: Buyer: Oh look! There's a **[STAIN]** on the green sweater….	*New-information focus*: 顾客: 哎, 你看! 你看这绿毛衣这块**[脏了]**。

2. Method

2.1. Participants

We tested 24 native speakers of Australian English (M_{age} = 21. 50 years; 21 females) and 24 native speakers of Mandarin Chinese (M_{age} = 27. 75 years; 20 females). All of the English speakers reported that they were born and raised in Australia, while the Mandarin speakers were born in Mainland China and had been living in an English-speaking country for less than nine years (M = 2. 75 years, range: 2 months–9 years). None reported any hearing or language impairment.

2.2. Materials

Different types of focus were manipulated largely based on Krifka's [1] proposals of the various pragmatic functions of focus. These include focus in response to wh-questions (wh-focus), focus used in correction statements (corrective focus), confirmation (confirmatory focus), and in parallel expressions (parallel focus). We also included focus that involves introduction of new information (new-information focus). Examples of the five focus types are illustrated in Table 1.

Dialogues written in casual English and Chinese were constructed to elicit participants' production of prosodic focus. In each language, we used two dialogues, where each dialogue contained pairs involving the same word tokens in a focused versus an unfocused realisation. For each of the focused and unfocused tokens, we measured 6 prosodic parameters: duration, mean F_0, maximum F_0, F_0 range, mean rms-intensity, and maximum rms-intensity. Different types of focus appeared throughout both dialogues, although not equally often. Unfocused tokens were defined as given information that had already been made salient by the focused tokens earlier in the dialogues. In each language, there were 21 pairs of focused and unfocused tokens, with 12 pairs in the first dialogue (2 wh-, 2 corrective, 2 confirmatory, 2 parallel, 4 new-information) and 9 pairs in the second dialogue (4 corrective, 1 confirmatory, 4 new-information). In consequence, we report a total of 12,096 measurements (2 languages X 21 pairs X 2 focus levels X 6 prosodic parameters X 24 participants).

The English and Mandarin dialogues were comparable in that each serves as a close translation of the other, with only small deviations (e.g., phonological relatives in the respective languages where the script has one of the actors making a hearing error). Another minor deviation in translation can be found in some adjectives (e.g., whether the colour of the sweater was "red" or "blue"), as we attempted to maintain similar levels of vowel frontness and/or openness. Apart from these minor variations, both sets of dialogues involved the focused and unfocused tokens within the same discourse contexts. To optimise comparability between the focused and unfocused tokens, we ensured that each focused token and its unfocused counterpart occurred in the same utterance position. Further, the utterance positions of the focused and unfocused tokens for most pairs were the same across both languages.

2.3. Procedures

All participants were individually tested in a sound-attenuated booth. Recordings were made using a headset microphone connected to a laptop via an audio interface. All dialogues were performed with the experimenter, who spoke in a lively style. The first dialogue involved a conversation between a buyer (the participant) and a street vendor (the experimenter). In the second dialogue, the participant played a high-school student who was being questioned by a police inspector (the experimenter). Recording sessions for each dialogue lasted for approximately five minutes, and both roles had equal number of turns (9 in the first dialogue, 11 in the second). Pairs of focused and unfocused tokens were deliberately excluded from the experimenter's portion of the dialogue.

Participants sat opposite the experimenter and were asked to read through each dialogue before each recording session. To ensure successful elicitation of focus, participants were encouraged to immerse themselves in their roles and be "as natural and genuine as possible". In addition, the experimenter asked all participants to pay careful attention to how they chose to speak in each dialogue. However, the experimenter gave no explicit instructions to emphasise the focused tokens. All participants were tested by the same English-Mandarin bilingual experimenter.

3. Results

All focused and unfocused word tokens were segmented and annotated based on simultaneous inspection of the waveform and the spectrogram in Praat [12], from which measurements for each prosodic parameter were extracted. For each dialogue, we conducted a series of 2-way 2 (language: English vs. Mandarin) X 2 (focus levels: focused vs. unfocused) mixed ANOVAs. Separate analyses were conducted to examine the focus production of each prosodic parameter for each focus type. Cross-language differences in English and Mandarin speakers' use of each prosodic parameter are revealed as significant interactions between language and focus levels. Significance threshold (at $\alpha = .05$) was adjusted using the Benjamini-Hochberg false discovery rate control procedure [13]. Since our interests are in cross-language differences, we report only the significant interactions. All significant cross-language differences observed in the first and second dialogues are illustrated in Figures 1 and 2 respectively.

3.1. First Dialogue

Analyses revealed a significant cross-language difference in the degree to which English and Mandarin speakers increased their duration for wh-focus, $F(1, 46) = 14.34$, $p < .001$. Simple effects of focus for the English and Mandarin speakers revealed that the increase in duration for wh-focus was longer in Mandarin ($p < .001$), although it was also significant in English speakers ($p = .001$). No other cross-language durational differences were found for the other focus types.

For F_0, there were significant cross-language differences in F_0 range only for wh-focus, $F(1, 46) = 7.56$, $p = .009$, and new-information focus, $F(1, 46) = 7.17$, $p = .01$. In wh-focus, only Mandarin speakers significantly expanded their F_0 range, ($p = .007$). In new-information focus, both speakers significantly expanded their F_0 range, with higher increase in speakers of Mandarin than English (both *p*-values $< .001$). There was also a cross-language difference in mean F_0 for new-information focus, $F(1, 46) = 7.53$, $p = .009$, where English speakers produced a higher increase than Mandarin speakers, (both *p*-values $< .001$). There were no significant cross-language differences in maximum F_0.

For mean and maximum rms-intensity, significant cross-language differences occurred only in the production of new-information focus, with both greater increases occurring in English rather than Mandarin (all *p*-values $< .001$).

3.2. Second Dialogue

There were no significant cross-language differences for duration, F_0 range, or any of the intensity measures. For mean F_0, results revealed a significant cross-language difference for new-information focus, $F(1, 46) = 9.30$, $p = .004$, such that Mandarin speakers produced greater increase ($p < .001$) than English speakers ($p = .001$). We also found a significant interaction for mean F_0 in corrective focus, $F(1, 46) = 23.77$, $p < .001$, in which Mandarin speakers showed greater increase ($p < .001$) than English speakers ($p = .014$). For maximum F_0, cross-language difference was found in production of corrective focus, $F(1, 46) = 15.27$, $p < .001$, where only Mandarin speakers showed the significant increase ($p < .001$). There were no significant cross-language differences for any of the other types of focus.

Figure 1. *Significant cross-language differences between English and Mandarin in the first dialogue (red = Mandarin, light blue = English)*

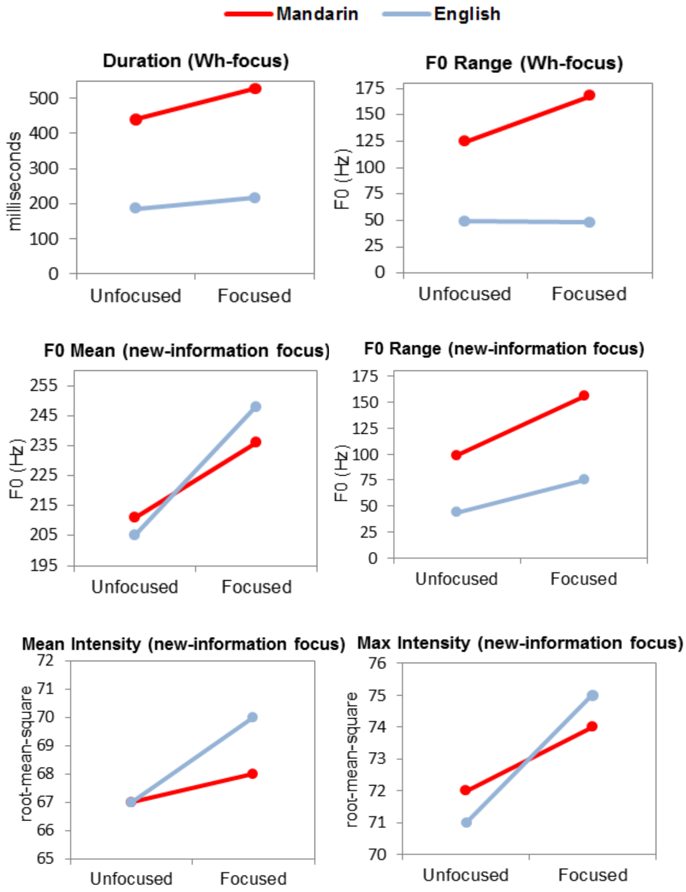
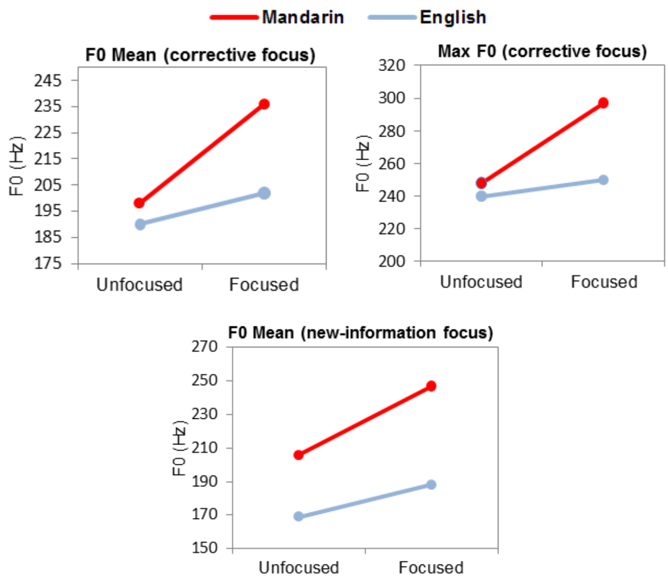


Figure 2. *Significant cross-language differences between English and Mandarin in the second dialogue (red = Mandarin, light blue = English)*



4. Discussion

The present experiment sheds new light both on the production of prosodic focus in general, as well as on the language-specific strategies that underlie speakers' use of prosody. In support of previous research [5, 6, 14], we show that native speakers of Mandarin resemble English speakers in their tendency to signal focus by manipulation of duration, pitch range, and intensity. However, extending prior work, we have further discovered instances where English and Mandarin speakers did not pattern similarly in the degree to which they employed the various prosodic parameters.

Two aspects of our findings are especially worthy of note. Firstly, whether and how speakers of Mandarin and English differ in their exploitation of each prosodic parameter depends on both the specific types of focus and the discourse-pragmatic contexts provided by each of the dialogues. For instance, in the first dialogue, cross-language differences occurred mostly in cases of new-information focus, where Mandarin speakers produced stronger increase in pitch range, while English speakers exerted greater increase in intensity. In the second dialogue, in contrast, cross-language differences occurred where Mandarin speakers produced new-information focus with greater increase in average and maximum pitch. Secondly, there were cases where cross-language differences occurred only for certain focus types and only in one of the dialogues, such as the greater increase in duration and pitch range for Mandarin wh-focus observed in the first dialogue only, and the cross-language differences in the production of corrective focus found only in the second dialogue. Therefore, how focus is prosodically expressed depends both on the specific discourse context and on the pragmatic function that focus serves in the utterance information structure.

This difference in our results for the different pragmatic expressions of focus has, as signalled in the introduction, potential implications for how focus is modeled in linguistic theory. Even though it may be most parsimonious to view focus as a unitary construct in information structure theory [8, 9], our findings suggest that speakers are more likely to realise focus in some pragmatic expressions than in others, and may also prefer their precise prosodic realisations of focus to differ from one pragmatic function to another.

The cross-dialogue differences that we observed suggest that discourse function also affects prosodic realisation of focus. In our first dialogue, the participant's role involved negotiation of a purchase; in the second, the role involved the report of a past event. The results showed that prosodic effects tended to be greater in the first dialogue, in which, had the situation been real, the speaker's money was at stake, than in the second, where there was little at stake for the witness; the differing findings may indeed indicate that our participants engaged enthusiastically in their role-playing task!

Apart from the variation across focus types and dialogues, the cross-language findings are intriguing because they indicate subtle variation in the use of the same prosodic resources that are available and used by speakers in both languages. The production of new-information focus in the domain of pitch gave the most reliable results. Across both dialogues, Mandarin speakers reliably show greater increase in pitch for new-information focus (either as pitch range or mean/maximum). For one thing, the fact that Mandarin had greater pitch increase than English is surprising because pitch in Mandarin also plays a crucial role in determining tone identity. One possible explanation for this cross-language difference in pitch could be that speakers of different languages vary in the level of attention they pay to each prosodic parameter when signaling focus. When speakers choose for some reason to speak carefully, they tend to modify their output in ways that are similar to prosodic focus (e.g., articulating more slowly and loudly [15]). Since pitch in Mandarin also serves another purpose, Mandarin speakers may need to pay more careful attention to pitch realisation to mark focus, so that the pitch information for lexical tones remains intact. This in turn may lead to more exaggerated use of pitch.

Of course, one potential counter-argument would be that Mandarin speakers already compensate for the dual roles of pitch by manipulating pitch range (as opposed to pitch shape). It is still an empirical question as to whether they exaggerate their increase in pitch on top of this strategy. A different explanation could be that Mandarin *listeners* would rely more on pitch information. From this point of view, Mandarin speakers would be more inclined to produce more exaggerated focal pitch because Mandarin listeners rely more on the pitch information than English speakers. To test this idea, future research could conduct a perceptual task where pitch information is rendered uninformative. For example, Cutler and Darwin [16] found that English speakers could still entrain to prosodic structure for locating sentence focus [17] even when pitch cues were removed by monotonising the sentences. A replication of this experimental paradigm in Mandarin could help address whether the greater increase in pitch observed in Mandarin speakers reflected a stronger perceptual reliance.

A final question that warrants further research is why the cross-language difference in pitch across the two dialogues was primarily observed in new-information focus. This is particularly interesting in that newness versus givenness is often cited as the classic distinction in utterance information structure and might thus be considered most likely to pattern similarly across languages.

From a methodological standpoint, we agree with Xu [18] that systematic experimental procedures are vital to fostering knowledge on language processing. The present experiment provides a novel approach in eliciting a more naturalistic form of speech that was nonetheless produced under controlled conditions. Furthermore, since participants were never instructed to emphasise any of the focused tokens, we argue that the prosodic focus elicited in the present experiment is a good reflection of speech production in natural settings.

5. Conclusion

Our findings provide evidence of language-specificity in prosodic processing where speakers' production of prosodic focus can differ even when the same prosodic resources are employed. At the same time, we present data showing how focus produced under various discourse-pragmatic contexts and dialogues can have quite different acoustic properties. The prosodic expression of focus may be more language-specific and more variable than previously thought.

6. Acknowledgements

Financial support was provided by the MARCS Institute and the ARC Centre of Excellence for the Dynamics of Language. We would like to thank Jason Shaw for his help and guidance. We also thank Ann Burchfield for her comments, and Steven Fazio and Mark Antoniou for technical support and advice.

7. References

[1] M. Krifka, "Basic notions of information structure," in *Interdisciplinary Studies of Information Structure 6*, C. Féry, G. Fanselow and M. Krifka (eds.). Potsdam: Universitätsverlag Potsdam, 2006, pp. 000-000.

[2] D. L. Bolinger, "Around the edge of language," *Harvard Educational Review*, vol. 34, pp. 282-296, July 1964.

[3] J. F. Hay, M. Sato, A. E. Coren, C. L. Moran and R. L. Diehl, "Enhanced contrast for vowels in utterance focus: A cross-langauge study," *Journal of the Acoustical Society of America,* vol. 119, pp. 3022-3033, May 2006.

[4] A. Rialland and S. Robert, "The intonational system of Wolof," *Linguistics*, vol. 39, pp. 893-939, September 2001.

[5] Y. Xu, "Effects of tone and focus on the formation and alignment of the f0 contour," *Journal of Phonetics*, vol. 27, pp. 55-105, January 1999.

[6] Y. Chen and C. Gussenhoven, "Emphasis and tonal implementation in Standard Chinese," *Journal of Phonetics*, vol. 36, pp. 724-746, June 2008.

[7] M. Sugahara, "Post-focus prosodic phrase boundaries in Tokyo Japanese: Asymmetric behavior of an F0 cue and domain-final lengthening," *Studia Linguistica*, vol. 59, pp. 144-173, October 2005.

[8] M. Rooth, "A theory f focus interpretation," *Natural Language Semantics*, vol. 1, pp. 75-116, February 1992.

[9] M. Krifka and R. Musan, "Information structure: Overview and linguistic issues," in *The Expression of Information Structure*, M. Krifka and R. Musan (eds.). Berlin: Mouton de Gruyter, 2012, pp. 1-44.

[10] I. C. Ouyang and E. Kaiser, "Prosody and information structure in a tone language: An investigation of Mandarin Chinese," *Language and Cognitive Processes*, May 2013.

[11] K. Ito, S. R. Speer and M. E. Beckman, "Information status and pitch accent distribution in spontaneous dialogues in English," in *Proceedings of Speech Prosody 2004*, pp. 279-282, 2004.

[12] P. Broesma, "Praat, a system for doing phonetics by computer," *Glot International*, vol. 5, pp. 341-345, December 2001.

[13] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 57, pp. 289-300, 1995.

[14] Y. Xu and C. X. Xu, "Phonetic realization of focus in English declarative intonation," *Journal of Phonetics*, vol. 33, pp. 159-157, April 2005.

[15] M. A. Picheny, N. I. Durlach and L. D. Braida, "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech," *Journal of Speech, Language and Hearing Research*, vol. 29, pp. 434-446, December 1986.

[16] A. Cutler and C. J. Darwin, "Phoneme-monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency," *Perception and Psychophysics*, vol. 29, pp. 217-224, May 1981.

[17] A. Cutler, "Phoneme-monitoring reaction time as a function of preceding intonation contour," *Perception and Psychophysics*, vol. 20, pp. 55-60, January 1976.

[18] Y. Xu, "In defense of lab speech," *Journal of Phonetics*, vol. 38, pp. 329-336, July 2010.